

---

# PROSITE: a dictionary of sites and patterns in proteins

---

**Amos Bairoch**

Department of Medical Biochemistry, University of Geneva, 1 rue Michel Servet, 1211 Geneva 4, Switzerland

---

## Background

PROSITE is a compilation of sites and patterns found in protein sequences. The use of protein sequence patterns (or motifs) to determine the function of proteins is becoming very rapidly one of the essential tools of sequence analysis. This reality has been recognized by many authors [1,2]. While there have been a number of recent reports [3,4] that review published patterns, no attempt had been made until very recently [5,6] to systematically collect biologically significant patterns or to discover new ones. It is for these reasons that we have developed, since 1988, a dictionary of sites and patterns which we call PROSITE.

Some of the patterns compiled in PROSITE have been published in the literature, but the majority have been developed, in the last two years, by the author.

## Format

The PROSITE database is composed of two ASCII (text) files. The first file (PROSITE.DAT) is a computer-readable file that contains all the information necessary to programs that make use of PROSITE to scan sequence(s) with pattern(s). This file also includes, for each of the patterns described, statistics on the number of hits obtained while scanning for that pattern in the SWISS-PROT protein sequence data bank [7]. Cross-references to the corresponding SWISS-PROT entries are also present in that file. The second file (PROSITE.DOC), which we call the textbook, contains textual information that documents each pattern. A user manual (PROUSER.TXT) is distributed with the database, it fully describes the format of both files. A sample textbook entry is shown in Figure 1 with the corresponding data from the pattern file.

## Leading concepts

The design of PROSITE follows four leading concepts:

**Completeness.** For such a compilation to be helpful in the determination of protein function, it is important that it contains a significant number of biologically meaningful patterns.

**High specificity of the patterns.** In the majority of cases we have chosen patterns that are specific enough not to detect too many unrelated sequences, yet that detect most if not all sequences that clearly belong to the set in consideration.

**Documentation.** Each of the patterns is fully documented; the documentation includes a concise description of the family of protein that it is supposed to detect as well as an explanation on the reasons that led to the selection of the particular pattern.

**Periodic reviewing.** It is important that each pattern be periodically reviewed, so as to insure that it is still relevant.

## Content of the current release

Release 6.10 of PROSITE (February 1991) contains 375

documentation entries describing 433 different patterns. The list of these entries is provided in Appendix 1.

## Distribution

PROSITE is distributed on magnetic tape and on CD-ROM by the EMBL Data Library. For all enquiries regarding the subscription and distribution of PROSITE one should contact:

EMBL Data Library  
European Molecular Biology Laboratory  
Postfach 10.2209, Meyerhofstrasse 1  
6900 Heidelberg, Germany  
Telephone: (+49 6221) 387 258  
Telefax : (+49 6221) 387 519 or 387 306  
Electronic network address: DATALIB@EMBL.BITNET

PROSITE can be obtained from the EMBL File Server [8]. Detailed instructions on how to make best use of this service, and in particular on how to obtain PROSITE, can be obtained by sending to the network address NETSERV@EMBL.BITNET the following message:

HELP  
HELP PROSITE

If you have access to a computer system linked to the Internet you can obtain PROSITE using FTP (File Transfer Protocol), from the following file servers:

GenBank On-line Service [9]  
Internet address: genbank.bio.net (134.172.1.160)

NCBI  
Internet address: ncbi.nlm.nih.gov (130.14.20.1)

The present distribution frequency is four releases per year. No restrictions are placed on use or redistribution of the data.

## REFERENCES

1. Doolittle R.F. (In) Of URFs and ORFs: a primer on how to analyze derived amino acid sequences., University Science Books, Mill Valley, California, (1986).
2. Lesk A.M. (In) Computational Molecular Biology, Lesk A.M., Ed., pp17-26, Oxford University Press, Oxford (1988).
3. Barker W.C., Hunt T.L., George D.G. Protein Seq. Data Anal. 1:363-373(1988).
4. Hodgman T.C. CABIOS 5:1-13(1989).
5. Bork P. FEBS Lett. 257:191-195(1989).
6. Smith H.O., Annau T.M., Chandrasegaran S. Proc. Natl. Acad. Sci. USA 87:826-830(1990).
7. Bairoch A., Boeckmann B. Nucleic Acids Res. 19:2247-2249(1991).
8. Stoehr P.J., Omund R.A. Nucleic Acids Res. 17:6763-6764(1989).
9. Benton D. Nucleic Acids Res. 18:1517-1520(1990).

## 1a) A documentation (textbook) entry

```
{PDOC00107}
{PS00116; DNA_POLYMERASE_B}
{BEGIN}
*****
* DNA polymerase family B signature *
*****
```

Replicative DNA polymerases (EC 2.7.7.7) are the key enzymes catalyzing the accurate replication of DNA. They require either a small RNA molecule or a protein as a primer for the de novo synthesis of a DNA chain. On the basis of sequence similarities a number of DNA polymerases have been grouped together [1 to 5] under the designation of DNA polymerase family B. The polymerases that belong to this family are:

- Human polymerase alpha.
- Yeast polymerase I, polymerase III, and polymerase REV3.
- Polymerases of viruses from the herpesviridae family (Herpes type I and II, Epstein-Barr, Cytomegalovirus, and Varicella-zoster).
- Polymerases from Adenoviruses.
- Polymerases from Baculoviruses.
- Vaccinia virus polymerase.
- Bacteriophage T4 polymerase.
- Podoviridae bacteriophages Phi-29, M2, and PZA polymerase.
- Tectiviridae bacteriophage PRD1 polymerase.
- Putative polymerases from yeast *K. lactis* linear plasmids pGKL1 and pGKL2.
- Putative polymerase from the maize mitochondrial plasmid-like S1 DNA.

Six regions of similarity (numbered from I to VI) are found in all or a subset of the above polymerases. The most conserved region (I) includes a perfectly conserved tetrapeptide which contains two aspartate residues. The function of this conserved region is not yet known, however it has been suggested [3] that it may be involved in binding a magnesium ion. We use this conserved region as a signature for this family of DNA polymerases.

- Consensus pattern: [YA]-x-D-T-D-S-[LIVMT]
- Sequences known to belong to this class detected by the pattern: ALL.
- Other sequence(s) detected in SWISS-PROT: chicken vitellogenin 2.
- Note: the residue in position 1 is Tyr in every family B polymerases, except in phage T4, where it is Ala.
- Last update: February 1991 / Text revised.

- [ 1] Jung G., Leavitt M.C., Hsieh J.-C., Ito J.  
Proc. Natl. Acad. Sci. U.S.A. 84:8287-8291(1987).
- [ 2] Bernad A., Zaballo A., Salas M., Blanco L.  
EMBO J. 6:4219-4225(1987).
- [ 3] Argos P.  
Nucleic Acids Res. 16:9909-9916(1988).
- [ 4] Wang T.S.-F., Wong S.W., Korn D.  
FASEB J. 3:14-21(1989).
- [ 5] Delarue M., Poch O., Todro N., Moras D., Argos P.  
Protein Engineering 3:461-467(1990).

```
{END}
```

## 1b) The corresponding entry in the pattern file

```
ID  DNA POLYMERASE_B; PATTERN.
AC  PS00116;
DT  APR-1990 (CREATED); NOV-1990 (DATA UPDATE); NOV-1990 (INFO UPDATE).
DE  DNA polymerase family B signature.
PA  [YA]-x-D-T-D-S-[LIVMT].
NR  /RELEASE=16,18364;
NR  /TOTAL=26(26); /POSITIVE=25(25); /UNKNOWN=0(0); /FALSE_POS=1(1); /FALSE_NEG=0(0);
CC  /TAXO-RANGE=?BEPV; /MAX-REPEAT=1;
DR  P09884, DPOASHUMAN, T; P13382, DPOASYEAST, T; P15436, DPODASYEAST, T;
DR  P14284, DPOXSYEAST, T; P03261, DPOLSADE02, T; P04495, DPOLSADE05, T;
DR  P05664, DPOLSADE07, T; P06538, DPOLSADE12, T; P04415, DPOLSBPT4, T;
DR  P08546, DPOLSHCMVA, T; P03198, DPOLSEBV, T; P04293, DPOLSHSV11, T;
DR  P07917, DPOLSHSV1A, T; P04292, DPOLSHSV1K, T; P09854, DPOLSHSV1S, T;
DR  P07918, DPOLSHSV21, T; P09252, DPOLSVZVD, T; P09804, DPO1SKLULA, T;
DR  P05468, DPO2SKLULA, T; P10582, DPOL$MAIZE, T; P18131, DPOL$NPVAC, T;
DR  P06856, DPOL$VACCV, T; P03680, DPOL$BPPH2, T; P06950, DPOL$BPPZA, T;
DR  P10479, DPOL$BPPRD, T;
DR  P02845, VIT2$CHICK, F;
DO  PDOC00107;
//
```

Figure 1. Sample data from PROSITE

**Appendix 1. List of patterns documentation entries in release 6.10 of PROSITE****Post-translational modifications**

N-glycosylation site  
 Glycosaminoglycan attachment site  
 Tyrosine sulfatation site  
 cAMP- and cGMP-dependent protein kinase phosphorylation site  
 Protein kinase C phosphorylation site  
 Casein kinase II phosphorylation site  
 Tyrosine kinase phosphorylation site  
 N-myristoylation site  
 Amidation site  
 Aspartic acid and asparagine hydroxylation site  
 Vitamin K-dependent carboxylation domain  
 Phosphopantetheine attachment site  
 Prokaryotic membrane lipoprotein lipid attachment site  
 Farnesyl group binding site (CAAX box)

**Domains**

Endoplasmic reticulum targeting sequence  
 Peroxisomal targeting sequence  
 Gram-positive cocci surface proteins 'anchoring' hexapeptide  
 Nuclear targeting sequence  
 Cell attachment sequence  
 ATP/GTP-binding site motif A (P-loop)  
 EF-hand calcium-binding domain  
 Actinin-type actin-binding domain signatures  
 Cofilin/tropomyosin-type actin-binding domain  
 Kringle domain signature  
 EGF-like domain cysteine pattern signature  
 Type II fibronectin collagen-binding domain  
 Hemopexin domain signature  
 'Trefoil' domain signature  
 Chitin recognition or binding domain signature  
 WAP-type 'four-disulfide core' domain signature

**DNA or RNA associated proteins**

'Homeobox' domain signature  
 'Homeobox' antennapedia-type protein signature  
 'Homeobox' engrailed-type protein signature  
 'Paired box' domain signature  
 'POU' domain signature  
 Zinc finger, C2H2 type, domain  
 Nuclear hormones receptors DNA-binding region signature  
 Eryf1-type zinc finger domain  
 Poly(ADP-ribose) polymerase zinc finger domain  
 Leucine zipper pattern  
 Fos/jun DNA-binding basic domain signature  
 Myb DNA-binding domain repeat signatures  
 Myc-type, 'helix-loop-helix' putative DNA-binding domain signature  
 p53 tumor antigen signature  
 'Cold-shock' DNA-binding domain signature  
 CTF/NF-I signature  
 ETS-domain signatures  
 SRF-type transcription factors DNA-binding and dimerization domain  
 Transcription factor TFIID repeat signature  
 eIF-4A family ATP-dependent helicases signatures  
 Eukaryotic putative RNA-binding region RNP-1 signature  
 Bacterial activator proteins, araC family signature  
 Bacterial activator proteins, crp family signature  
 Bacterial activator proteins, gntR family signature  
 Bacterial activator proteins, lacI family signature  
 Bacterial activator proteins, lysR family signature  
 Bacterial histone-like DNA-binding proteins signature  
 Histone H2A signature  
 Histone H2B signature  
 Histone H3 signature  
 Histone H4 signature  
 HMG1/2 signature  
 HMG-I and HMG-Y DNA-binding domain (A+T-hook)  
 HMG14 and HMG17 signature  
 Protamine P1 signature  
 Ribosomal protein L5 signature  
 Ribosomal protein L11 signature

Ribosomal protein L14 signature  
 Ribosomal protein L23 signature  
 Ribosomal protein L39/L46 signature  
 Ribosomal protein S7 signature  
 Ribosomal protein S8 signature  
 Ribosomal protein S9 signature  
 Ribosomal protein S10 signature  
 Ribosomal protein S11 signature  
 Ribosomal protein S12 signature  
 Ribosomal protein S15 signature  
 Ribosomal protein S17 signature  
 Ribosomal protein S18 signature  
 Ribosomal protein S19 signature  
 DNA mismatch repair proteins mutL / hexB / PMS1 signature

**Enzymes****Oxidoreductases**

Zinc-containing alcohol dehydrogenases signature  
 Iron-containing alcohol dehydrogenases signature  
 Insect-type alcohol dehydrogenase / ribitol dehydrogenase family signature  
 Aldo/keto reductase family signatures  
 L-lactate dehydrogenase active site  
 Glycerate-type 2-hydroxyacid dehydrogenases signature  
 Hydroxymethylglutaryl-coenzyme A reductases signatures  
 3-hydroxyacyl-CoA dehydrogenase signature  
 Malate dehydrogenase active site signature  
 Malic enzymes signature  
 Glucose-6-phosphate dehydrogenase active site  
 Bacterial quinoprotein dehydrogenases signatures  
 Aldehyde dehydrogenases active site  
 Glyceraldehyde 3-phosphate dehydrogenase active site  
 Acyl-CoA dehydrogenases signatures  
 Glutamate dehydrogenases active site  
 Dihydrofolate reductase signature  
 Pyridine nucleotide-disulphide oxidoreductases active site  
 Nitrite reductases and sulfite reductase putative siroheme-binding sites  
 Uricase signature  
 Cytochrome c oxidase subunit I, copper B binding region signature  
 Cytochrome c oxidase subunit II, copper A binding region signature  
 Multicopper oxidases signatures  
 Lipoygenases, putative iron-binding region signature  
 Extradiol ring-cleavage dioxygenases signature  
 Intradiol ring-cleavage dioxygenases signature  
 Bioppterin-dependent aromatic amino acid hydroxylases signature  
 Copper type II, ascorbate-dependent monooxygenases signatures  
 Cytochrome P450 cysteine heme-iron ligand signature  
 Copper/Zinc superoxide dismutase signature  
 Manganese and iron superoxide dismutases signature  
 Ribonucleotide reductase large subunit signature  
 Ribonucleotide reductase small subunit signature  
 Nitrogenases component 1 alpha and beta subunits signature

**Transferases**

Thymidylate synthase active site  
 Methylated-DNA--protein-cysteine methyltransferase active site  
 N-6 Adenine-specific DNA methylases signature  
 N-4 cytosine-specific DNA methylases signature  
 C-5 cytosine-specific DNA methylases signatures  
 Serine hydroxymethyltransferase pyridoxal-phosphate attachment site  
 Phosphoribosylglycinamide formyltransferase active site  
 Aspartate and ornithine carbamoyltransferases signature  
 Thiolases signatures  
 Chloramphenicol acetyltransferase active site  
 cysE / lacA / nodL acetyltransferases signature  
 Phosphorylase pyridoxal-phosphate attachment site  
 UDP-glucuronosyl and UDP-glucosyl transferases signature  
 Purine/pyrimidine phosphoribosyl transferases signature  
 S-Adenosylmethionine synthetase signatures  
 EPSP synthase active site  
 Aspartate aminotransferases pyridoxal-phosphate attachment site  
 Hexokinases signature  
 Galactokinase signature

Phosphofructokinase signature  
 Protein kinases signatures  
 Pyruvate kinase active site signature  
 Phosphoglycerate kinase signature  
 Aspartokinase signature  
 ATP:guanido phosphotransferases active site  
 PTS proteins phosphorylation sites signatures  
 Adenylate kinase signature  
 Phosphoribosyl pyrophosphate synthetase signature  
 Eukaryotic RNA polymerase II heptapeptide repeat  
 DNA polymerase family B signature  
 Galactose-1-phosphate uridyl transferase active site signature  
 CDP-alcohol phosphatidyltransferases signature  
 Rhodanese active site

#### **Hydrolases**

Phospholipase A2 active sites signatures  
 Lipases, serine active site  
 Colipase signature  
 Carboxylesterases type-B active site  
 Alkaline phosphatase active site  
 Fructose-1-6-bisphosphatase active site  
 Serine/threonine specific protein phosphatases signature  
 Tyrosine specific protein phosphatases active site  
 Prokaryotic zinc-dependent phospholipase C signature  
 3'/5'-cyclic nucleotide phosphodiesterases signature  
 Sulfatases signature  
 Pancreatic ribonuclease family signature  
 Alpha-lactalbumin / lysozyme C signature  
 Lysosomal alpha-glucosidase / sucrase-isomaltase active site  
 Alpha-L-fucosidase putative active site  
 Uracil-DNA glycosylase signature  
 Serine carboxypeptidases, serine active site  
 Zinc carboxypeptidases, zinc-binding regions signatures  
 Serine proteases, trypsin family, active sites  
 Serine proteases, subtilisin family, active sites  
 ClpP proteases active sites  
 Eukaryotic thiol (cysteine) proteases active site  
 Ubiquitin carboxyl-terminal hydrolase, putative active-site signature  
 Eukaryotic aspartyl proteases active site  
 Neutral zinc metalloproteases, zinc-binding region signature  
 Insulin-degrading enzyme / E.coli protease III signature  
 recA signature  
 Proteasome subunits signature  
 Asparaginase / glutaminase signature  
 Urease active site  
 Beta-lactamases classes -A, -C, and -D active site  
 Arginase and agmatinase signatures  
 Inorganic pyrophosphatase signatures  
 Acylphosphatase signatures  
 ATP synthase alpha and beta subunits signature  
 ATP synthase gamma subunit signature  
 ATP synthase delta (OSCP) subunit signature  
 E1-E2 ATPases phosphorylation site  
 Sodium and potassium ATPases beta subunits signatures  
 Cutinase, serine active site

#### **Lyases**

DDC / GAD / HDC pyridoxal-phosphate attachment site  
 Orotidine 5'-phosphate decarboxylase signature  
 Phosphoenolpyruvate carboxylase active site  
 Ribulose biphosphate carboxylase large chain active site  
 Fructose-bisphosphate aldolase active site  
 KDPG and KHG aldolases active site signatures  
 Isocitrate lyase signature  
 DNA photolyases signature  
 Carbonic anhydrases signature  
 Fumarate lyases signature  
 Enolase signature  
 Serine/threonine dehydratases pyridoxal-phosphate attachment site  
 Enoyl-CoA hydratase signature  
 Tryptophan synthase alpha chain signature  
 Tryptophan synthase beta chain pyridoxal-phosphate attachment site  
 Delta-aminolevulinic acid dehydratase active site

#### **Isomerases**

Alanine racemase pyridoxal-phosphate attachment site  
 Peptidyl-prolyl cis-trans isomerase signature  
 Triosephosphate isomerase active site  
 Xylose isomerase signatures  
 Phosphoglucose isomerase signature  
 Phosphoglycerate mutase family phosphohistidine signature  
 Eukaryotic DNA topoisomerase I active site  
 Prokaryotic DNA topoisomerase I active site  
 DNA topoisomerase II signature

#### **Ligases**

Aminoacyl-transfer RNA synthetases class-I signature  
 Aminoacyl-transfer RNA synthetases class-II signatures  
 ATP-citrate lyase and succinyl-CoA ligases active site  
 Glutamine synthetase signatures  
 Ubiquitin-conjugating enzymes active site  
 Phosphoribosylglycinamide synthetase signature  
 ATP-dependent DNA ligase putative active site

#### **Others enzymes**

Isopenicillin N synthetase signatures  
 Site-specific recombinases signatures  
 Thiamine pyrophosphate enzymes signature  
 Biotin-requiring enzymes attachment site  
 2-oxo acid dehydrogenases acyltransferase component lipoyl binding site

#### **Electron transport proteins**

Cytochrome c family heme-binding site signature  
 Cytochrome b5 family, heme-binding domain signature  
 Cytochrome b/b6 signatures  
 Thioredoxin family active site  
 Glutaredoxin active site  
 Type-1 copper (blue) proteins signature  
 2Fe-2S ferredoxins, iron-sulfur binding region signature  
 4Fe-4S ferredoxins, iron-sulfur binding region signature  
 Rieske iron-sulfur protein signatures  
 Flavodoxin signature  
 Rubredoxin signature

#### **Other transport proteins**

Class I metallothioneins signature  
 Ferritin iron-binding region signature  
 Transferrins signatures  
 Plant hemoglobins signature  
 Arthropod hemocyanins / insect LSPs signatures  
 ATP-binding proteins 'active transport' family signature  
 Binding-protein-dependent transport systems inner membrane component signature  
 Serum albumin family signature  
 Lipocalins signature  
 Cytosolic fatty-acid binding proteins signature  
 LBP / BPI / CETP family signature  
 Uteroglobin family signatures  
 Mitochondrial energy transfer proteins signature  
 Sugar transport proteins signatures  
 Prokaryotic sulfate-binding proteins signature  
 Amino acid permeases signature  
 Anion exchangers family signatures  
 MIP / Nodulin-26 / glpF family signature  
 Insulin-like growth factor binding proteins signature

#### **Structural proteins**

43 Kd postsynaptic protein signature  
 Actins signatures  
 Annexins phospholipid/calcium-binding domain signature  
 Clathrin light chain signature  
 Connexins signatures  
 Crystallins beta and gamma 'Greek key' motif signature  
 Dynamin family signature  
 Intermediate filaments signature  
 Kinesin motor domain signature  
 Neuromodulin (GAP-43) signatures  
 Profilin signature

Surfactant associated polypeptide SP-C palmytoylation sites  
 Synapsins signatures  
 Synaptobrevin signature  
 Tropomyosins signature  
 Tubulin subunits alpha and beta signature  
 Tubulin-beta mRNA autoregulation signal  
 Tau and MAP2 proteins repeated region signature  
 Neuraxin and MAP1B proteins repeated region signature  
 F-actin capping protein beta subunit signature  
 Amyloidogenic glycoprotein signatures  
 Cadherins extracellular repeated domain signature  
 Insect larval cuticle proteins signature  
 Gas vesicles protein GVPa signature  
 Gas vesicles protein GVPc repeated domain signature  
 NMePhe pili methylation site  
 Potexviruses and carlaviruses coat protein signature

#### Receptors

Neurotransmitter-gated ion-channels signature  
 G-protein coupled receptors signature  
 Visual pigments (opsins) retinal binding site  
 Bacterial rhodopsins retinal binding site  
 Receptor tyrosine kinase class II signature  
 Receptor tyrosine kinase class III signature  
 Growth factor and cytokines receptors family signature  
 Integrins alpha chain signature  
 Integrins beta chain cysteine-rich domain signature  
 Photosynthetic reaction center proteins signature  
 Photosystem I psaA and psaB proteins signature  
 Phytochrome chromophore attachment site  
 Speract receptor repeated domain signature  
 TonB-dependent receptor proteins signatures  
 Type-II membrane antigens family signature

#### Cytokines and growth factors

Int-1 family signature  
 HBGF/FGF family signature  
 Nerve growth factor family signature  
 Platelet-derived growth factor (PDGF) family signature  
 TGF-beta family signature  
 TNF family signature  
 Interferon alpha and beta family signature  
 Interleukin-1 signature  
 Interleukin-2 signature  
 Interleukin-6 / G-CSF / MGF signature  
 Interleukin-7 signature

#### Hormones and active peptides

Adipokinetic hormone family signature  
 Bombesin-like peptides family signature  
 Calcitonin / CGRP / IAPP family signature  
 Chromogranins / secretogranins signatures  
 Gastrin / cholecystokinin family signature  
 Glucagon / GIP / secretin / VIP family signature  
 Glycoprotein hormones beta chain signature  
 Insulin family signature  
 Natriuretic peptides signature  
 Neurohypophysial hormones signature  
 Pancreatic hormone family signature  
 Parathyroid hormone family signature  
 Somatotropin, prolactin and related hormones signatures

Tachykinin family signature  
 Cecropin family signature  
 Mammalian defensins signature  
 Insect defensins signature  
 Endothelins / sarafotoxins signature

#### Toxins

Plant thionins signature  
 Snake toxins signature  
 Heat-stable enterotoxins signature  
 Aerolysin type toxins signature  
 Shiga/ricin ribosomal inactivating toxins active site signature  
 Channel forming colicins signature  
 Staphylococcal enterotoxins / Streptococcal pyrogenic exotoxins signatures  
 Membrane attack complex components / perforin signature

#### Inhibitors

Pancreatic trypsin inhibitor (Kunitz) family signature  
 Bowman-Birk serine protease inhibitors family signature  
 Kazal serine protease inhibitors family signature  
 Soybean trypsin inhibitor (Kunitz) protease inhibitors family signature  
 Serpins signature  
 Potato inhibitor I family signature  
 Squash family of serine protease inhibitors signature  
 Cysteine proteases inhibitors signature  
 Tissue inhibitors of metalloproteinases signature  
 Cereal trypsin/alpha-amylase inhibitors family signature  
 Disintegrins signature

#### Others

Pentraxin family signature  
 Immunoglobulins and major histocompatibility complex proteins signature  
 Prion protein signature  
 Cyclin signature  
 Proliferating cell nuclear antigen signature  
 Arrestins signature  
 Chaperonins signature  
 Heat shock hsp70 proteins family signatures  
 Heat shock hsp90 proteins family signature  
 Ubiquitin signature  
 SRP54-type proteins GTP-binding domain signature  
 GTP-binding elongation factors signature  
 Eukaryotic initiation factor 4D hypusine signature  
 S-100/ICaBP type calcium binding protein signature  
 Hemolysin-type putative calcium-binding region signature  
 Small, acid-soluble spore proteins, alpha/beta type, signature  
 Caseins alpha/beta signature  
 Legume lectins signatures  
 Vertebrate galactoside-binding lectin signature  
 Lysosome-associated membrane glycoproteins signatures  
 Glycophorin A signature  
 Seminal vesicle protein I repeats signature  
 HCP repeats signature  
 Bacterial ice-nucleation proteins octamer repeat  
 Cell cycle proteins ftsW / rodA / spoVE signature  
 Staphylocoagulase repeat signature  
 11-S plant seed storage proteins signature  
 Dehydrins signature  
 Small hydrophilic plant seed proteins signature  
 Thaumatin family signature